VOLUME 63 WHOLE NO. 306
NUMBER 11 1949

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

A Comparison of Two Tests of Intelligence Administered to Adults

By

ANNA S. ELONEN

Division of Psychiatry
University of Chicago Clinics

Accepted for publication, July 13, 1949

Price $1.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1950, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION

TABLE OF CONTENTS

I. Introduction 1

II. Selection of Subjects 4

III. Procedure 6

IV. Presentation of Results 8

A. Results for the Adult Group 8

B. Results for the Student Group 9

C. Results for the Feebleminded Group 10

D. Results for the Patients with Convulsive Disorders 10

E. Results for the Patients with Organic Disease of the Brain 11

F. Results for the Patients with Emotional Disturbance 11

G. The Nature of the Tests in Relation to the Results 13

V. Discussion of the Speed, Accuracy, and Variability Scores 18

A. The Findings on the Speed Score 18

B. The Findings on the Accuracy Score 19

C. Comparison of the Speed and Accuracy Scores 19

D. Comparison of the Speed and Accuracy Scores with Performance 20

E. The Findings on the Variability Score 21

VI. Implications 23

A. Diagnostic Significance of IQ Differences between the Stanford-Binet and the Kuhlmann Tests 23

B. Diagnostic Significance of the Speed, Accuracy, and Variability Scores 24

C. Implications for Further Study 25

VII. Summary and Conclusions 27

A. A Summary of the Agreement between the Hypotheses and the Results 27

1. Primary Hypotheses and Results 27

2. Secondary Hypotheses and Results 29

B. Conclusions 30

Bibliography 34

CHAPTER I

INTRODUCTION[1]

THE psychological literature abounds with reference to the limitations of the Revised Stanford-Binet Scale and the Kuhlmann Tests of Mental Development as measures of adult intelligence (13, 19, 20, 58). In spite of the fact that these tests were not designed primarily for adults, many clinics have continued to use either one or both of these tests as individual measures of adult intelligence for the value of their diagnostic patterns. Unfortunately, few clinicians have published their findings with respect to the internal patterns which they have uncovered and continue to use in differential diagnosis. In the opinion of many clinicians, the Revised Stanford-Binet as well as the Kuhlmann tests have rich and varied possibilities for comprehensive interpretation, which to date have been claimed chiefly for the projective methods. If the standard intelligence tests were to receive the same careful scrutiny, the seeking out of the internal patterns, and the qualitative evaluation of results, the present blanket acceptance of the numerical score as the only aim of intelligence testing would cease to be the prevalent practice (45).

The far-seeing choice of items, the breadth of functions explored in either the Stanford-Binet or the Kuhlmann alone, and the marked contrast between the types of items in the Stanford-Binet and the Kuhlmann cause some examiners to prefer the joint use of these measures to the popular Wechsler-Bellevue Scale. It remains to be seen how useful the unpublished Jastak scale will be (24), as well as the less well known test by Cattell (9). It is felt that the current enthusiasm for the ever-growing body of projective techniques and for the latest measures of intelligence have tended to push aside unjustly the older established measures which, in the light of the added research findings from both the psychiatric and the neurological fields, have proved to be unusually sensitive and comprehensive (54, 6).

A comparison of the Stanford-Binet and the Kuhlmann tests on adults was undertaken, because comparative knowledge about the two tests was felt to be of specific value in different types of clinical situations. Frequently it is necessary to obtain measures on the same individuals within short intervals. Demands for re-examination are made by medical and legal practice and by research method. The effects of induced fever or drugs upon performance, the difference in abilities before and after surgery on the brain, and progress during psychotherapy are measured by psychological tests. Although the medical profession is chiefly interested in the test results

[1] I wish to thank all the people through whose help this study was made possible. I am especially indebted to John and Eileen Kyle Marshall for their help in beginning the project, to Ruth French Liotta, John Brozenec, and Katherine Hamilton, for the numerous contacts they made possible, to all our staff and especially to Ruth Lambert Bromberg, Mary Onken Styrt, Ruth L. Mohns, and Adelaide Black, to Dr. Warren C. Murray and Mrs. Anna P. Johnson of the Dixon State Hospital, to Dr. Agnes Sharp and the late Dr. David B. Rotman of the Municipal Court for their cooperation in permitting a number of their patients to be tested, and to Dr. David Slight, Dr. Ward C. Halstead, Dr. Frank H. Finch, the late Dr. Andrew Brown, Dr. Charlotte Babcock, Dr. Lester MacLean, Dr. Frances Swineford, Dr. Douglas N. Buchanan, Dr. Ralph W. Tyler, Dr. Stephen M. Corey, and Dr. Mandel Sherman for their suggestions and criticisms, to Marianne Scheuer Zimmt and Carol P. Smith, secretaries, for their willing help, and finally to all the people who consented to take the tests.

per se, the psychologist, on the other hand, is concerned not only with the individual patient but also with the comparison of obtained results. Results of tests given to patients with the same types of diagnoses may be compared in order to find similarities and differences in the patterns between widely different types of cases. Observations of this kind can be made only through individual study, and at present there are available for this purpose few individual intelligence tests for which comparative observations have been made.

In order to provide data concerning the relationship between the Revised Stanford-Binet and the Kuhlmann Tests of Mental Development, a comparison of the results of these two tests on different groups of adults was begun in 1938 as a research project of the Clinical Psychological Service of the Department of Psychiatry in the University of Chicago Clinics. These measures had been selected for clinical use because they provided unusually adequate internal patterns, diagnostic leads, and complementary pictures of the individuals tested.

A review of the literature reveals that only a few comparisons between the two scales have been attempted (7, 8, 60). The literature is, however, filled with attacks on the inadequacies of adult intelligence tests in general. One of the repeated criticisms has been that existing intelligence tests have not taken into consideration the obvious fact that the intelligence of children and of adults differs in quality rather than in amount (3, 10, 19, 52). On the other hand, qualitative interpretation of the results of such tests is frequently considered an aid to diagnosis in clinical practice.

The inclusion of the speed factor in tests administered to adults, and the relation of speed to age and deterioration has been a controversial point in the literature (12, 25, 38, 39). The Kuhlmann scale, which gives a separate measure of speed, affords an opportunity for further investigation of this factor.

The original aim of the present study was to compare the performance of the same individual on the Revised Stanford-Binet and the Kuhlmann Tests of Mental Development, but later on it was decided that it would be desirable to compare the results in one clinical group with the results in another. In order to obtain a representative survey of the various clinical groups, the following diagnostic classifications were selected as chief clinical diagnostic categories: a) the convulsive state, b) organic disease of the brain, c) emotional disturbance, and d) feeblemindedness. Since the examinations were to be made in a clinical setting, the pathological groups were obtained readily from the different clinic services. These pathological groups were to be compared with groups of normal adults and students.

One would predict a high degree of correspondence between two adequately standardized measures, such as the Revised Stanford-Binet and the Kuhlmann, when applied to normal adults. On the other hand, clinical experience suggests that a different relationship between these scales will be found in test results of normal adults and those of individuals who show either organic or psychogenic pathology. The Revised Stanford-Binet, because of its few timed tests, the generally familiar nature of the material, and the element of rapport between the examiner and the patient should not reflect changes in functioning ability as readily as the strictly timed Kuhlmann and its unfamiliar and abstract material

and more impersonal relationship. It is generally known that both the organic and the psychogenic cases are slower in their reactions, and that patients with organic disturbance depend upon remote rather than recent learning. Both types of patients, those with emotional disturbance and those with organic involvement, can be better motivated, the closer the interaction permitted between the individuals in the test situation.

Therefore, the primary hypotheses to be tested by means of this study are: (*a*) that the correlation between the two tests in normal adult groups is high, and (*b*) that the amount of agreement decreases in proportion to the severity of the pathological involvement. Furthermore, the patients with organic involvement should show less agreement between test results than the patients suffering from neurotic and psychotic disorders, unless marked deterioration exists in the psychotic individual.

The new measures of speed and accuracy introduced by the Kuhlmann tests (28) should prove to be diagnostically significant, for the speed with which an individual functions should be found to be lowered appreciably in the pathological cases, and the degree to which the speed has been reduced should depend upon the degree to which the pathology has invaded the system. One would expect that the accuracy score would not be affected to as great an extent as the speed factor. The speed and accuracy scores should not, as a rule, be lowered as greatly in the emotionally disturbed patients as in those suffering from organic insult. In the same way, one would hypothesize that the accuracy scores would not be as seriously affected by pathology as the speed scores. These scores in the feebleminded group cannot be compared, for the results fall below the lower limits of scoring. Because of the students' emphasis on speed, which is well known, one would expect that the speed score of the student group would be the highest achieved by any group. The significance of the difference between the means of the speed and the means of the accuracy scores from group to group will be established.

Although to date no work has been done with the variability score, this measure should prove to be a useful indicator in differentiating the various diagnostic categories. Because of the lability of the emotionally disturbed patient, one would set up the hypothesis that the variability scores of this group would be greater than those of any other group. A significant relation between the intelligence level and the variability score would also be expected from the nature of a measure of this kind. The statistical reliability of the difference between the means of the variability scores on the various groups will be computed. The possible relationship between variability and many other factors will be plotted and inspected to see what relationship exists in order that more exact information on the measure may be available. One would also expect that this measure might show definite relationship with such factors as speed, accuracy, and sex.

CHAPTER II

SELECTION OF SUBJECTS

THE original aim of this study was to obtain measures on the Stanford-Binet and the Kuhlmann tests for the same adults, who were to be divided into the following diagnostic classifications: a) normal adults, b) students, c) feeble-minded adults, d) patients with the diagnosis of convulsive state, e) patients with organic disease of the brain, and f) patients with emotional disturbance.

a. Normal adults. In the normal adult group an attempt was made to get as representative a cross-section of occupational groups as possible. The final distribution according to the Beckman rating scale (4) can be seen in Table 1. The distribution was slightly skewed in the direction of superiority of occupational status. None of the common rating scales included housewives; therefore, these were placed in a separate category. Because all the testing was done in urban areas, no farm group was included.

TABLE 1

DISTRIBUTION OF THE NORMAL ADULT GROUP ACCORDING TO THE BECKMAN OCCUPATIONAL SCALE

Occupation	Number of Cases
Unskilled Manual	34
Semi-Skilled	28
Skilled Manual	79
Sub-Professional	61
Professional	29
Housewife	20
Total	251

In the group labeled normal adults were included all the adults on whom there was no information or history of an abnormality in behavior, nor any observation during the test situation which would lead one to question this opinion. The group totaled 251, of whom 131 were women, 120, men. They ranged in age from 18 through 62 years, the median age for the women being 30, for the men, 28. Variation in schooling was from the third grade level through the Ph.D. level. The median school grade for the entire group, and for the men and women separately, was the twelfth. All were white, born in the United States and had been out of school at least two years.

These cases were obtained from numerous sources. The largest number were from the many settlement houses in the Chicago area. Frequently, entire club groups consented to take the tests. Another source was one of the large department stores which permitted a group of its personnel to be tested. This group consisted mainly of floorwalkers and stenographers. Within the hospital itself, some employees and some convalescent patients consented to take the tests. Through individuals in one of the university departments and in two boarding houses for young business and professional women, practically all members of these groups were tested. The remainder were single contacts. In fact, a surprisingly large number were friends of those who had had the tests and, in turn, persuaded others to volunteer.

Some examiners feel that it is extremely difficult to persuade the adult person to be tested (11). However, in our experience with at least 15,000 cases, refusals have been almost nonexistent, in spite of the fact that the individuals represented almost all ethnic groups, in all conceivable home and institutional environments, and included some of the most irritable patients suffering from organic brain damage. This explanation is given to forestall any possible criticism that any group of volunteers might not be a representative sample. In view of this experience, it is felt that the fact that the subjects were volunteers in no way made them a less representative sample.

In testing this group, it was observed that the results on those who were attending school appeared to differ from those of other adults (10). Therefore, it was considered desirable to form a separate group made up entirely of those attending school at the time of the testing (see *b* below).

b. Students. The students were obtained for the greater part from the class in medical psychology at the University of Chicago School of Medicine, and, therefore, most of them were men and medical students. The remainder of the men and most of the women came from other divisions of the University of Chicago, although there were small groups from other colleges. The total number of this group was 204, of whom 103 were men and 101 women.

In school placement they ranged from college freshmen to those about to receive doctoral degrees. The fifth year of college was the median for the men, and the third year of college for the women. In age they ranged from 17 years through 30, the median for the men being 23, for the women, 21.

c. Feebleminded adults. The third group was made up of people on whom a definite diagnosis of feeblemindedness had been made. The greatest portion of these were inmates of a state institution for the feebleminded; some were seen at the Psychiatric Institute at court; some were referred by social agencies; and the others were patients in the University of Chicago Clinics who had been diagnosed as feebleminded. The total number of this group was 91, of whom 61 were women and 30 were men. The age range was from 16 through 49 years, with the median age for the women at 29 and for the men at 26.

d. Patients with diagnosis of convulsive state. Patients subject to convulsive seizures formed a fourth group (Table 2). These individuals were seen in the neurological clinic. The group included 51 men and women, in a 26-25 distribution, with an age range from 16 through 45, a median age of 24. Because of the small number, this group was not separated according to sex.

TABLE 2

DIAGNOSIS OF PATIENTS WITH CONVULSIVE DISORDERS

Diagnosis	Number of Cases
Convulsions major, cryptogenic	44
Convulsions major, cryptogenic; convulsions minor, cryptogenic	6
Convulsions major, cryptogenic; convulsions minor, cryptogenic; hysteria	1
Total	51

e. Patients with organic disease of the brain. The group diagnosed as having organic disease of the brain was composed of patients referred by the neurosurgical and neurological clinics. Diagnoses for these cases came chiefly under the headings of tumor, disease, and trauma. The total number of this group was only 43, 28 men and 15 women. Consequently, no attempt was made to divide the group. The age range was from 16 through 53 years, with the median at 26.

f. Patients with emotional disturbance. The last group was made up of emotionally disturbed patients who had been seen and diagnosed by the psychiatrists in the in- and outpatient clinics. None of the patients was particularly deteriorated or as much out of contact with reality as the patients usually seen in some psychopathic wards. The total number of cases (adults and students) was 113. Among the nonstudents, the number of men was 42; women, 43; the total, 85, with the median age at 26. The median age of the women was 27, the ages ranging from 16 through 40 years. The median for the men was 24.5 years, with a range from 16 through 58.

The majority of the 85 adult cases could be classed as having some type of neurosis, including maladjustments and a few psychopathic personalities, while there was a smaller number of psychoses. The neurotic group included 66 patients, ranging in age from 16 through 58.

In the same way that it had been found practicable to divide the normal adult group into those who no longer attended school and those who were students, patients with emotional disturbance who were still attending school were separated from the main functional group. There were 28 students, ranging in age from 16 through 25.

Chapter III

PROCEDURE

The normal adult and student groups were given the tests in the same manner. One portion of each group was given the L form of the Stanford first and the Kuhlmann second; in another section, this order of the tests was reversed; in a third, the M form of the Stanford was given first, the Kuhlmann second; and in the fourth, this order was reversed. At first, only one test was given at a time, but because of moving, travel convenience, and other reasons, some tests could not be completed. As a result, in the majority of cases, both tests were given the same day with a rest period between the two tests. The tests were given at the Clinics, settlement houses, or private homes under the usual individual testing conditions.

All but 35 of the original tests were given by one examiner, and the remainder were given by one other person. Both examiners had had a great many years of experience in administering individual tests, and had been examiners on the Kuhlmann standardization. In the cases of the patients diagnosed as having organic disease of the brain or emotional disturbance or convulsive state, a few tests were given by other members of the psychological staff of the Clinics, and the L form of the Stanford-Binet was usually given first.

The administration of the Revised Stanford-Binet conformed with the usual procedures (56, pp. 60-68). According to customary clinical practice, all three superior levels were given when the basal was established above the first superior level. The Kuhlmann was given in accordance with the standard methods (28, pp. 104-112).

Kuhlmann states that growth curve values may be converted into mental ages and then directly into IQ's (28). However, since absurdly high IQ's result when the mental age equivalents reach 45 years and over, he suggests that if IQ's must be given for adults, the P.A. (per cent of average) score (28, pp. 98-99) should be used to find the probable IQ at the age of ten. He estimates that for the brightest adults the maximum IQ's obtained in this way would probably be about 180. The extent to which he underestimated this maximum can be seen from the fact that IQ's as high as 285 were obtained by this formula.

As this illustrates, the necessary lowering of the IQ's at the upper levels could not be accomplished by dropping to the probable IQ at the age of ten. Consequently, some other method was sought to make the cut which seemed to be clinically acceptable as well as logical in relation to results of other tests.

Kuhlmann gave a mental age of 45 for Heinis unit 429 and conceded that some performances would be scored well over 500. This is fairly accurate, since the highest total Heinis score obtained was 494, with individual items still higher. Although the mental age for Heinis unit 429 is already so high, Kuhlmann nevertheless proceeded to give credit for Heinis units through 528. The estimated mental age for this would reach astronomical figures, well beyond common or clinical sense. The Heinis units used at these levels contribute to the unusual results. For these reasons and also because the highest Heinis units seem to be overweighted for speed, an experiment was tried of using 429 as the maximal score. This involved

the rescoring of tests 79-89 inclusive. Empirically, this cut seemed to be too much, partly because of the effect of the cylindrical scoring (28, p. 125) on the size of the cut for some items. Eventually, these eleven tests were cut so that they were scored at four levels instead of six; that is, no test received credit beyond 462 mental units.

The tests were first scored by the examiner and then rescored by two other examiners. When very unusual responses were obtained for which the examples given by Terman and Merrill did not suffice, the authors were consulted and the consensus of other clinical psychologists was sought. Kuhlmann scoring was carefully watched for clerical errors; all results were checked and scored by two examiners.

CHAPTER IV

PRESENTATION OF RESULTS

A. RESULTS FOR THE ADULT GROUP ($N = 251$)

As has been explained, the tests were originally given using both the L and the M forms of the Revised Stanford-Binet, some given before and some after the Kuhlmann, so that four different categories result: M_1, M_2, L_1, L_2. Variations in the means, standard deviations, and correlations among the subgroups can probably be explained on the basis of the sampling and the small size of the groups.

The differences between the various subgroups were not sufficiently great to prevent the combining of them to form total groups of men and women, and of both together (Table 3). None of the differences between the correlations for the subgroups or men's and women's groups was statistically reliable.

When the men were grouped together (Table 3) and the Kuhlmann was scored on the conventional 16/10[2] basis, a correlation was found between the Stanford-Binet and Kuhlmann of $.819 \pm .030$. The correlation of $.774 \pm .035$ for the women was only slightly lower. For the entire 251 men and women the correlation between the two tests was $.797 \pm .023$ (Table 3).[3] Cutting the scores for speed (16/10 cut) as previously described (page 7) raised the correlation to $.841 \pm .018$.[4] This differed sig-

TABLE 3

STATISTICAL FINDINGS FOR NORMAL ADULT GROUP
(Revised Stanford-Binet, Forms L and M, and Kuhlmann Developmental Scale)

	Men					Women				
	N	Mean	σ	r	σr	N	Mean	σ	r	σr
	Total					Total				
S-B	120	117.88	19.90	.819	.030	131	119.27	22.27	.774	.035
K 16/10		102.93	25.20				110.77	33.54		

Men and Women Combined

	Total					Range		
	N	Mean	σ	r	σr	Men	Women	Combined
S-B	251	119.04	20.80	.797	.023	70–149	71–149	70–149
K 16/10		107.20	30.10			64–166	59–285	59–285
K 16/10 cut		103.50	23.50	.841	.018	64–149	59–166	59–166

[2] 16/10 designates Kuhlmann's scoring system in which the IQ of an adult is calculated as it probably would have been at the age of ten.

[3] To see if any differences existed between correlations using mental ages or IQ's as measures, correlations using the mental age were made on certain groups as samples; the results were essentially the same.

[4] Because of the disproportionately large number of college graduates in this sampling the two tests were correlated on a smaller group in which the number of college graduates was reduced to the per cent found in the populations of the states of Iowa, West Virginia, and Illinois as reported by Warren S. Thompson. The relationship was not appreciably affected, for the correlation obtained for this group was $.798 \pm .024$ (52).

TABLE 4

STATISTICAL FINDINGS FOR STUDENT GROUP
(Revised Stanford-Binet, Forms L and M, and Kuhlmann Developmental Scale)

	Men					Women				
	N	Mean	σ	r	σr	*N*	Mean	σ	r	σr
	Total					Total				
S-B	103	138.40	8.04	.515	.072	101	134.45	11.67	.604	.063
K 16/10		147.29	29.55				140.40	32.83		

Men and Women Combined

	Total					Range		
	N	Mean	σ	r	σr	Men	Women	Combined
S-B	204	136.40	10.70	.541	.049	108–152	86–152	86–152
K 16/10		146.40	31.80			94–258	74–285	74–285
K 16/10 cut		129.50	16.85	.625	.043	94–160	74–160	74–160

nificantly beyond the one per cent level from the correlation obtained by use of uncut scores. The group thus became more homogeneous, as indicated by the decreased standard deviation, and the correlation was raised because the unduly high scores which overweighted one end of the distribution had been reduced. This change in range from 59 through 285 to 59 through 166 illustrates the shifting of the high scores closer to the median.

When the difference in IQ between the Stanford-Binet and the Kuhlmann on the same individual was considered, regardless of sign, the mean difference in IQ points between the Stanford-Binet and the Kuhlmann (16/10 cut) *without regard to sign,* was ± 16.84. One-fourth of the group did not show a difference greater than 9 points. In most cases the Stanford-Binet was the higher, as indicated by the means of the two tests. One possible explanation of the consistent difference in favor of the Stanford-Binet is that no one in the group had attended school within the last two years. The average person who has been away from an educational situation for any period of time finds difficulty in working under pressure of time, as required by the Kuhlmann (14). In the second place, the more personal relationship between the subject and the examiner on the Stanford-Binet probably appeals more to the average person and may add greater motivation. The Stanford-Binet is also less abstract and less monotonously presented. Finally, in some of the Stanford-Binet items, one can depend on past learning to a greater extent than one can in the Kuhlmann.

B. RESULTS FOR THE STUDENT GROUP ($N = 204$)

The results of the tests of the students are shown in Table 4. As can be seen, the differences were again small enough to be explained on the grounds of sampling. In general, the correlations were markedly lower than those for the normal adults, since the students were a much more homogeneous group.

On the whole, comparing these results with another Chicago study, it was found that the standard deviation of the

Stanford-Binet of the total group and subgroups of men were close to or less than the standard deviation of 9.9 obtained by this other study (2). However, the standard deviations of the total group, the total group of women, as well as two of the subgroups for women was greater.

The differences in correlation which appeared between the M form of the Stanford-Binet and the Kuhlmann and the L form and the Kuhlmann were not statistically reliable. They can probably be explained chiefly by the small size of the samples, the use of the uncut scores, and the low ceiling possible on the Stanford-Binet in comparison to the Kuhlmann. Inspection of the groups showed no definite clues to the reason.

When the men students were grouped together (Table 4) and the Kuhlmann was scored on the usual 16/10 basis, the correlation was .515 ± .072. The correlation of .604 ± .063 for the women was somewhat higher. For the entire 204 men and women students, the correlation between the two tests was .541 ± .049. When the cuts were made, the correlation was raised to .625 ± .043, again a statistically reliable difference. As in the case of the normal adult group, the student group thus became more homogeneous, as indicated by the decreased standard deviations, and the correlation was raised because the numerous unduly high scores on the Kuhlmann no longer existed. The upper end of the range dropped from 285 to 160, which was more reasonable but actually perhaps somewhat low for the calibre of this group.

The mean difference in IQ points between the Stanford-Binet and the Kuhlmann (16/10 cut), without regard to sign, was ±11.53. Almost half of the group did not have a difference greater than 9 points. As indicated by the means of the two tests, the uncut Kuhlmann tended to be higher, but the difference was not as consistent as it was in the reverse direction for the normal adults. The cut Kuhlmann tended to be lower. Many of the students, of course, achieved a maximal rating on the Stanford-Binet.

C. Results for the Feebleminded Group ($N = 91$)

Since the median chronological age for the mentally defective group was higher than that of most feebleminded groups studied, the results tended to bring out what was to be expected from these studies; that is, that in groups of individuals of low intelligence level, the more reliable results are obtained from the younger age groups (8, 46).

The groups were all given the L form of the Stanford-Binet first. When the men and women were grouped together, the correlation for the total 91 was .770 ± .043 (Table 5). None of the scores for this group was high enough to be affected by the suggested cut. The mean difference in IQ for the whole group, without regard to sign, was ±7.56. The Stanford-Binet tended to be higher but was less consistently so for the women than for the men. It is difficult to state just how the tests were influenced, but the institutionalized women were much less cooperative in the test situation than the men. The factor of delinquency also played a greater part in the institutionalization of women than men. These results, therefore, would confirm the relationship between the two tests reported by Carlton in his study of defectives (8).

D. Results for the Patients with Convulsive Disorders ($N = 51$)

There was a statistically reliable difference between the means of the Stan-

TABLE 5

STATISTICAL FINDINGS FOR THE FEEBLEMINDED ADULT GROUP, FOR PATIENTS WITH CONVULSIVE STATES, AND FOR PATIENTS WITH ORGANIC DISEASE OF THE BRAIN (Revised Stanford-Binet, Form L, and Kuhlmann Developmental Scale, Uncut)

	Men and Women Combined					Range		
	N	Mean	σ	*r*	σr	Men	Women	Combined
Feebleminded Adults								
S-B	91	61.69	9.01	.770	.043	50–82	42–79	42–82
K 16/10		55.09	7.37			45–73	38–68	38–73
Patients with Convulsive States								
S-B	51	99.79	23.12	.736	.064			54–140
K 16/10		80.38	19.82					44–122
Patients with Organic Disease of the Brain								
S-B	43	100.08	18.84	.696	.079			62–143
K 16/10		74.04	15.09					52–118

ford-Binet and the Kuhlmann in the convulsive group. The standard deviations for both tests were also in the neighborhood of 20 points, and the correlation remained fairly high, .736 ± .064 (Table 5). The mean difference in IQ, without regard to sign, was ±21.12; this is almost as large as that for the total emotionally disturbed, and larger than for any other total group except the organic. The direction of the change was quite consistently towards a lower Kuhlmann. Only fifteen per cent of the cases showed a difference in IQ of 9 points or less. The only three cuts for speed were too small to alter the distribution.

E. RESULTS FOR THE PATIENTS WITH ORGANIC DISEASE OF THE BRAIN ($N = 43$)

The patients with organic disease of the brain also showed a statistically reliable difference between the means of the two tests. From the mean of 100.08 for the Stanford-Binet, the average rating dropped to 74.04 on the Kuhlmann. The group, therefore, seemed to be closer to the average ability of the general population than that reviewed by Hebb (18), since most of these cases tended to cluster within the average range on the Stanford-Binet. The difference was apparently fairly consistent, for the correlation was .696 ± .079 (Table 5). The mean difference in IQ points between the two tests on the same individual, without regard to sign, was ±26.30, the largest median difference to be observed. Less than 5 per cent of the group differed by not more than 9 points in IQ. Whereas less than 2 per cent of the normal adult and student populations had shown differences greater than 39 points, 20.9 per cent of the patients with organic disease of the brain showed as great or greater a discrepancy between the two tests. The only cut for speed was small, and did not affect the correlation.

F. RESULTS FOR THE PATIENTS WITH EMOTIONAL DISTURBANCE ($N = 113$)

Although the number of the entire emotionally disturbed group was not

large, it was felt to be desirable to differentiate the group in terms of sex, diagnosis, and school attendance. Therefore, the functional group was divided, after being considered as a complete unit, into adult and student groups, the criteria for the student group being the same as they had been for the unselected student group. It was also divided on the basis of neurotic and psychotic diagnostic classifications.

The adult group totaled 85, 66 of whom were diagnosed as neurotic, and 19 as psychotic. The distribution according to sex for this group was 42-43. The entire student group numbered 28.

The majority of the functional cases belonged in the L_1 group. The standard deviations were most like those for the normal adult group (see Tables 3 and 6). The Stanford-Binet means without exception were greater than those for the Kuhlmann test. The correlation of the Stanford-Binet and the uncut Kuhlmann for the entire adult functional group was .637 ± .064. The correlation for the cut Kuhlmann scores was .756 ± .046.

When the adult emotionally disturbed group was divided into those diagnosed as neurotic and those diagnosed as psychotic, the following correlations for the neurotic group were obtained. The cut changed the correlation only slightly. The correlation for the neurotic group when uncut was .780 ± .048; with the cut, it was .783 ± .032 (Table 6).

TABLE 6

STATISTICAL FINDINGS FOR PATIENTS WITH EMOTIONAL DISTURBANCE
(Revised Stanford-Binet, Form L, and Kuhlmann Developmental Scale, Uncut and Cut)

	Men and Women Combined					Range		
	N	Mean	σ	r	σr	Men	Women	Total
Adult Neurotics								
S-B	66	120.71	21.45	.780	.048			66–152
K 16/10 uncut		102.08	29.44					55–208
K 16/10 cut		98.89	23.30	.783	.032			55–154
Entire Adult Emotionally Disturbed								
S-B	85	120.74	20.06	.637	.064	83–152	66–149	66–152
K 16/10 uncut		102.50	33.00			56–259	55–166	55–259
K 16/10 cut		98.85	23.49	.756	.046	56–160	55–149	55–160
Combined Adults and Students								
S-B	113	121.23	20.63	.704	.047	83–152	66–152	66–152
K 16/10 uncut		106.27	34.96			56–259	55–196	55–259
K 16/10 cut		101.58	24.91	.777	.037	56–160	55–160	55–160

When the students[5] and adults were considered as one group the correlation for the uncut scores was .704 ± .047, and .777 ± .037 for the cut. When this entire group was divided by sex, the correlation for the men was .648 ± .078 (cut,

[5] Because of the small number included in the student functional group no further subdivisions were considered feasible. The correlation for this group on the uncut scores of the Kuhlmann was .791 ± .076; on the cut it was raised to .849 ± .053, a statistically reliable difference at the one per cent level of confidence.

.747 ± .059); for the women, .770 ± .054 (cut, .813 ± .045).

It was found that dividing the entire group of emotionally disturbed into the smaller subgroups of student and non-student, men and women, or psychotic and neurotic did not bring out any statistically reliable differences in the correlations except in the case of the psychotic and neurotic adults, in which a statistically reliable difference almost at the one per cent level of confidence was found in the correlation between the uncut Kuhlmann and the Revised Stanford-Binet.

In general, the mean differences in IQ points were large, more comparable with the differences for the organic pathological cases than for the normal adults or students. The one exception occurred in the case of the emotionally disturbed students whose mean difference was higher than that of the unselected student group but somewhat lower than that of the normal adult group. The mean difference for the adult group, without regard to sign, was ±23.65; for the adult neurotic group, ±23.36; for the student group, ±15.57; and for the entire emotionally disturbed group ±21.65. The direction of the shift reflected the fact that the Stanford-Binet scores were generally the higher. In summary, the differences between the two tests were much larger for the emotionally disturbed than for the unselected adults, whereas the differences between the tests for the students with emotional difficulties remained more like those of the normal students.

G. The Nature of the Tests in Relation to the Results

In general, the means for the various groups reach the levels which one might expect for the particular selection involved (Figure 1). For example, the means of the Stanford-Binet for the students are very high, while those of the normal adults and the emotionally disturbed cases are somewhat lower. However, the means for these two groups are surprisingly similar. The means for the Kuhlmann are consistently lower than the means for the Stanford-Binet. Both the normal adult group and the group of emotionally disturbed patients consist largely of brighter than average individuals, as brought out by the mean of the Stanford-Binet. The clinical impression was that the normal adult group was somewhat above average, a fact confirmed by the occupational groupings which were weighted toward the higher levels.

Whether the Stanford-Binet is a valid estimate of adult intelligence, and especially of superior intelligence, is certainly open to question. Undoubtedly, it is a much better measure for those individuals whose test results do not utilize the upper levels of the test. It is regrettable that the Stanford-Binet abruptly becomes such an ineffectual instrument at the upper levels when it has tapped so many diversified areas and has provided such an adequate diagnostic measure at the lower levels. At the Superior Adult levels, vocabulary and rote recall are emphasized disproportionately. Many functions are tapped through the Average Adult level, but cease to be measured beyond this stage.

The Kuhlmann, on the other hand, may give spuriously low results for those who rarely read from the moment compulsory instruction ceases. In addition, for those who have been out of school, the work under pressure of time and the quick shifts of types of subtests may also tend to lower the Kuhlmann rating. Un-

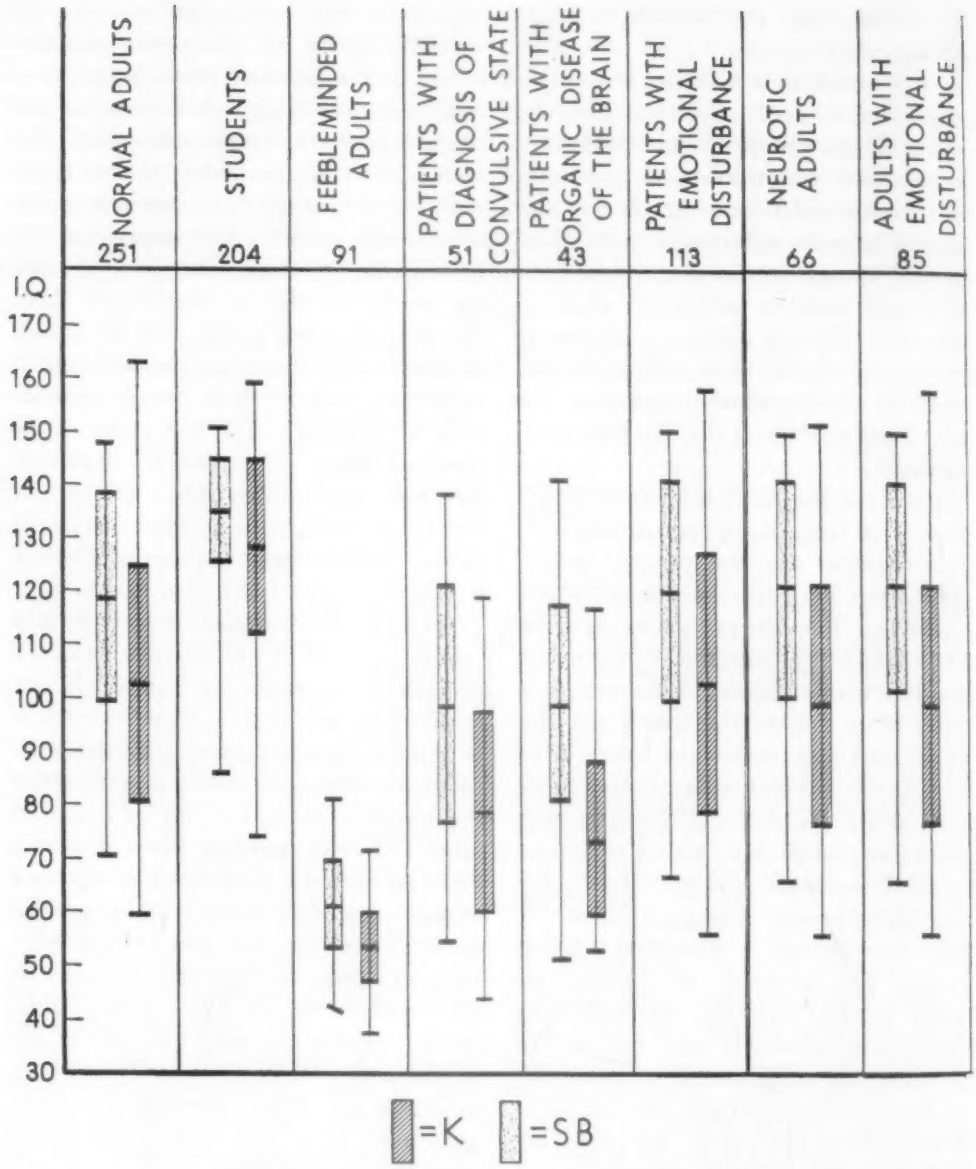

FIG. 1. The Ranges, Means, and Standard Deviations on the Stanford-Binet and the Kuhlmann (Uncut) Scales for all Groups.

doubtedly, this explains why at least some of the normal adults tended to have lower Kuhlmann scores. Another reason may exist in the inequivalency of standardization of the two tests.

One of the criticisms of the Stanford-Binet type of test has come from those who have worked with patients who have had operations on the brain (16). That the Stanford-Binet IQ can remain practically unaltered after the removal of entire lobes of the brain may perhaps imply an insensitivity of the test beyond certain limits. It must, of course, be re-

membered that although the IQ on the Stanford-Binet may not change significantly, markedly different internal patterns may appear if certain year levels are included in testing. For instance, in some types of organic disorders, the spatial tests are very significant, but if those year levels were not included in the testing of a particular individual, then the resulting patterns will not be diagnostically helpful. Also, the type of organic disorder makes a great deal of difference in the type of pattern obtained. The frontal lobe case, on the whole, usually has a very different type of pattern from that of the occipital case, and both depend on the amount of organic involvement.

In contrast, a lowering of functional ability, be it due to organic or functional causes, is reflected in the groups presented in this study in an appreciable lowering of the Kuhlmann IQ, as well as the speed and accuracy scores. The greater capacity of the Kuhlmann to show changes may be due to the nature of the test itself and to its effect on the individual tested. The test has a greater variety of items and provides a very short period for execution. The individual must have the capacity to pick out relevant points as well as the ability to shift his set constantly. It has been put forth that an individual is capable of productive thinking when he is able not only to work out relationships, either in the abstract or concrete areas, but also is sufficiently flexible to continue shifting his hypotheses until a satisfactory solution is achieved (59). Therefore, it would seem that the constant reorientation of the subject's attitude to new items in the spiral omnibus type of test is not a defect as indicated by Cattell (10), but rather, a marked asset, as suggested by these results.

The standard deviation of the normal adult group on the Revised Stanford-Binet was 20.8. This was higher than those obtained by Terman on children, but certainly is in line with the more recent studies on adults which indicate that the standard deviations probably would be in the neighborhood of 20 points (10, 49).

The standard deviations for the Kuhlmann tend to be greater than those for the Stanford-Binet in the normal adult, student, and emotionally disturbed groups. This one would expect from the greater range possible on the Kuhlmann. In the feebleminded, where the group is very homogeneous and the possible range of performance on either test is limited, the standard deviations of both tests are approximately equal. The reverse tendency, however, occurs in the patients with organic involvement, for in both the organic and the convulsive groups the standard deviations of the Stanford-Binet exceed those of the Kuhlmann. The consistent slowing down of reactions is sufficient to produce a marked constrictive effect on the range of the Kuhlmann scores. Also, the performance on the Stanford-Binet is greatly variable, some types of tests having withstood the onslaught of the disease and other types having been markedly affected. On the other hand, the Kuhlmann tests, depending less upon past learning, are affected much more consistently.

Dependence upon various types of past learning in the items of general intelligence tests has been criticized. Hartson concluded from his research that memory for the techniques used in arithmetical computations dims from lack of practice (17). McFarland's research also indicated that the recency of having had mathematical practice affected the re-

sults (36). The Kuhlmann relies less upon items involving learned arithmetical processes than probably any other general intelligence test. This should be considered a strength of the Kuhlmann, for so many studies have used tests which were as much as one-fourth mathematical in content, from which they have concluded that intelligence score curves reach a high point at 18 years of age and then decline (37). One cannot but question whether this peak was reached at this particular age because the tests had been given at the most optimal time for recalling learned mathematical processes.

All of the correlations obtained are statistically reliable and most of them are of high order. As might be expected from the homogeneity of the groups, the lowest correlation is found for the group of students. The emotionally disturbed student group, followed closely by the normal adult group, show the highest relationship between the Revised Stanford-Binet and the Kuhlmann.

The statistical reliability of the differences between the correlation of the diagnostic groups was computed, and very few statistically reliable differences between the correlations of the groups emerged. The only clear-cut differences were established between the students and the other categories, with the exception of the organic group. The statistical reliability at the one per cent level of confidence of the difference between the correlations of the normal adults and the students justifies the separation which was made on empirical grounds, and, therefore, emphasizes the dangers of using only students in standardization groups (10).

Although the correlations between the two tests were high for all groups, as brought out above, the reason was found in the consistency of the difference between the tests rather than in identity of the obtained results. With the exception of the student group and the feebleminded, the Kuhlmann rating tended to average 20 points below that of the Stanford-Binet, and the consistency of this relationship can be seen from comparisons of the mean of the two tests.

When the differences between the means of the two tests in each group were tested, the statistical reliability of the difference in each group was found to be well in excess of the one per cent level. Even where the difference between the means was only seven points, the significance of the difference was at the one per cent level of confidence.

The differences between the means of the Revised Stanford-Binet and the Kuhlmann showed an equal amount of statistical reliability between the normal and the pathological groups. Since the significant differences between the groups could not be discerned statistically from the comparison of the two means, the differences between the means of the Stanford-Binet from group to group and the differences between the means of the Kuhlmann for all groups were established. When this was done, statistically reliable differences at the one per cent level of confidence appeared between the same groups on the Revised Stanford-Binet as were found on the Kuhlmann. The only nonreliable relationships on either test occurred between the emotionally disturbed groups and the normal adults; the organic and the convulsive groups; and the subdivisions of the emotionally disturbed group.

Since the comparisons of the two means within each group and the comparison of each mean from group to group did not reveal any clear-cut differences between the various diagnostic classifications, the number of IQ points by which the Kuhlmann and the Revised Stanford-Binet test differed in each individual record was computed regardless of sign. The statistical reliability of the difference between groups of the mean differences in IQ points could then be tested. The normal adults showed a statistically reliable difference at the one per cent level of confidence between all other groups except the emotionally disturbed students and the convulsives. The normal adult and the convulsive

groups showed a statistically reliable difference in IQ points at the five per cent level of confidence.

With the student group, essentially the same differences in statistical reliability in IQ points emerged, the only exception being that the convulsive group showed a statistically reliable difference at the one per cent level of confidence rather than at the five per cent level. This group, like the normal adult group, failed to show any statistical reliability in IQ point differences from the student emotionally disturbed group.

When the organic group was compared with the other categories, less reliable differences were observed. In four cases, the normal adults, the students, the feebleminded, and the emotionally disturbed students, there was a statistically reliable difference at the one per cent level of confidence. No statistical reliability in the IQ point differences appeared between the organic group and the convulsive or the total emotionally disturbed group. Because of the organic pathology represented in both the patients with convulsive states and those with focal lesions, one was not surprised that no statistically reliable differences in IQ were revealed. The lack of differentiation between the emotionally disturbed group and the organic pathological groups is more difficult to explain but it may be accounted for by the emotional overlay which is a part of the picture of any organic condition. Since one manner in which most patients with organic involvement control their decreased efficiency is by compulsive defenses, this, coupled with the fact that a large number of the patients diagnosed as neurotic showed compulsive characteristics, may also in part account for this lack of differentiation between the two groups.

As one would have expected, the pattern of the mean differences in IQ points between the convulsive group and the other classifications followed that of the organic group almost identically. The only exceptions occurred in the comparison of the convulsive group with the normal adults, and the emotionally disturbed students, where the statistical reliability of the difference dropped to the five per cent level of confidence. The group of convulsive patients was outstanding for the reason that it presented so many who continued to function at an average level. For this reason, undoubtedly, the differences between this group and the normal adults were not more clear-cut.

The mean difference in IQ points between the feebleminded group and those of the others showed statistically reliable differences beyond the one per cent level for all groups.

When the mean differences in IQ points between the entire emotionally disturbed group were compared with the normal adults, the students, the feebleminded, and the functional student group, reliable differences beyond the one per cent level of confidence emerged. The lack of reliable differentiation between the convulsive and organic groups, and the emotionally disturbed, has already been discussed. As would have been expected, no statistically reliable differences occurred when the entire group of emotionally disturbed were compared with the adult functional group or the neurotic adult subdivision. This similarity is not surprising in view of the fact that the psychotic patients included in the emotional adult group were an extremely small percentage of the total, and that the adults formed a much larger percentage of the entire emotionally disturbed group than did the students.

When the differences of the means of the Kuhlmann, of the Revised Stanford-Binet, and of the IQ point difference between the two tests were compared, it was apparent that in a surprisingly large percentage of comparisons the Kuhlmann alone or the Stanford-Binet alone differentiated significantly between the groups, as did the mean difference in IQ points which combined the results of the two tests. The following combinations of groups were equally well differentiated by either test alone or by the difference in IQ points between the tests: the students and the normal adults, the organic pathological cases and the functional student cases, the organic pathological groups and the feebleminded group, the normal adults and the feebleminded, and all pathological classifications and the feebleminded. No statistically reliable difference occurred between the convulsive and the organic groups, and between the total functional group and its adult and adult neurotic components in any of the three comparisons.

The means of the Kuhlmann and the means of the Stanford-Binet differentiated reliably between the organic pathological cases and the total functional group, and between the students and the emotionally disturbed students. The mean differences in IQ points between the two tests failed to show any statistically reliable differentiations when these comparisons were made.

On the other hand, the means of the Kuhlmann and of the Stanford-Binet failed to show any significant difference between the normal adults and the emotionally disturbed group, and between the functional student cases and the entire functional group and its adult subdivisions, whereas the difference between the mean differences in IQ points did show statistically reliable variation.

Chapter V

DISCUSSION OF THE SPEED, ACCURACY, AND VARIABILITY SCORES

A. The Findings on the Speed Score

Definite tendencies appeared between the different groups with respect to the speed scores (Table 7). A difference, statistically reliable at the one per cent level of confidence, was found between the mean speed score of the normal adults and those of the students, the organic group and the convulsive group. On the other hand, no statistically reliable differences were found between the mean speed scores of the normal adults and the emotionally disturbed group or its adult or student subdivisions.

The fact that the emotionally disturbed group was made up of more individuals with neurotic than psychotic diagnoses and the fact that the emotionally disturbed group was of high calibre would tend to uphold the finding that there were no appreciable differences in the speed with which they worked.

No numerical comparisons, of course, could be made with the speed score of the feebleminded with any other group because only about one-fourth of the feebleminded individuals achieved speed within scorable limits.

The mean speed score of the student group differed reliably beyond the one per cent level of confidence from the mean speed scores of all the other groups.

It was interesting to note that the mean speed scores of the organic groups differed reliably at the one per cent level of confidence from the speed scores of all other groups except the convulsive group. Clinically, one would not expect a significant difference between the speed of the organic and the convulsive groups.

The mean speed score of the convulsive group showed statistically reliable differences from the mean speed scores of all groups except the organic group, which has already been brought out.

The entire emotionally disturbed group, as well as the adult and student subdivisions, differed reliably with regard to speed from all other major groups except the normal adults. This lack of differentiation from the normal adult group may suggest that, in so far as speed is an indication of functioning,

TABLE 7

Mean and Median Speed and Accuracy Scores on the Kuhlmann Scale

Diagnostic Classification	Speed		Accuracy	
	Median	Mean	Median	Mean
Normal adults	102.00	101.41	97.00	97.65
Students	114.00	114.50	104.00	103.38
Feeble-minded adults	Minimal	Minimal	Minimal	Minimal
Patients with diagnosis of convulsive state	85.00	88.06	95.00	95.24
Patients with organic disease of the brain	85.00	86.31	93.00	93.20
Patients with emotional disturbance				
Adults	94.00	104.42	101.00	98.90
Students	105.50	105.50	102.00	100.28
Total	97.00	99.00	101.00	99.30

the normal adult group, as any unselected adult group, may have possessed certain unstable tendencies.

The marked differences between these groups in the median speed scores reaffirm the fact brought out by previous investigators, among them McFarland, who stated that speed of reactions is one of the most important factors in individual differences in ability to react to mental tasks (36). According to these results, the speed of reaction may be influenced by functional as well as by organic factors.

B. The Findings on the Accuracy Score

In general, the accuracy score shows very much the same general pattern observed in the speed score on the different groups. The mean accuracy score of the normal adult group differed reliably at the one per cent level of confidence only between the students and between the organic group. At the five per cent level of confidence, there was a statistically reliable difference between the accuracy scores of the normal adult group and the convulsive and the entire emotionally disturbed group and the student division of the emotionally disturbed. Interestingly enough, the only group which did not show a statistically reliable difference from the normal adults in the accuracy score was the adult emotionally disturbed group. Again, in accuracy as in speed, the normal students differed significantly from every other group. The accuracy score followed the patterns of the speed score in the organic group in that the only nonreliable relationship was found between the convulsive group.

The emotionally disturbed group showed statistically reliable differences at the one per cent level of confidence in the relationship of the accuracy score between the different groups, with the exception of the already above-mentioned normal adult group. The adult emotionally disturbed group differed reliably from the students and from the groups showing organic pathology, but failed to differ reliably from the entire emotionally disturbed group and its student subdivision, and the already mentioned normal adult group. The student emotionally disturbed group showed statistically reliable differences from all groups except the entire group of emotionally disturbed and its adult subdivisions.

C. Comparison of the Speed and Accuracy Scores

The speed and accuracy scores among the students, both unselected and disturbed, were definitely much higher than those found in any other group. In addition, the speed score was definitely higher than the accuracy score among the students, although the accuracy was also above 100. For the normal adults, the speed score was slightly above norms while the accuracy was somewhat below norms; actually, both scores were close to the 103 mark. The opposite tendency appeared in the convulsive and organic cases, for the speed in most cases was appreciably lower than the accuracy. There was less difference between the speed and accuracy scores in the entire emotionally disturbed group than in any other.

In general, the median scores tended to give a more adequate picture of the different groups since the few deviant scores did not obscure the general tendency of the speed score to lag behind the accuracy score in the pathological

cases, both the organic and psychogenic. There was no marked discrepancy between the mean and median measures in any group except the adult emotionally disturbed, where the median fell 10 points below the mean. The median speed score showed a marked downward trend while the accuracy remained fairly normal for this group.

The median accuracy score of the emotionally disturbed group was somewhat higher than that of the normal adult group. One factor which may have been responsible for this was the fact that those with compulsive tendencies may have raised the average for the adult functional group as a whole. The cases with organic pathology differed from the functional group in that the accuracy rating often fell below average, and the speed score dropped to a phenomenally low point. In these cases, the accuracy and speed were both down, with the speed score much the lower. In the feebleminded group, both speed and accuracy were likely to fall still lower than those of any other group; in fact, the scores usually fell below the measurable level.

The maximum accuracy score possible on the Kuhlmann is much lower than the maximum speed score, and, therefore, because of the limited number of test items, the accuracy score cannot ever reach the high level achieved by the speed score.

Therefore, the findings support the hypothesis that the speed with which an individual functions is lowered appreciably in the pathological cases; and, in general, those suffering from organic pathology have a lower speed score than the psychogenic cases. As was predicted, the accuracy score was not affected to as great an extent as the speed factor. Without question, the speed of the student group, and also of the students suffering from emotional disturbance, was markedly higher than the median speed score of any of the other groups.

In general, the speed and accuracy scores of the clinical groups showed very similar patterns in the statistically reliable differences between the means. There was a very reliable difference between all the groups in the speed score, with the exception that the speed score of the organic and convulsive groups, and the emotionally disturbed patients and normal adults did not show any significant differentiation. In the case of the accuracy score, the same non-reliability emerged between the organic and convulsive groups.

D. Comparison of the Speed and Accuracy Scores with Performance

Because of the numerous references in the literature stating that age influences speed of performance, these two factors were correlated for the normal adult group. The correlation between the speed scores and age for the entire range of this group of 251 adults was $-.020 \pm .063$. When the data were checked for non-linearity, the correlation ratio of $.282 \pm .058$ revealed a significant trend towards a negative relationship between age and speed. Although the younger individuals in the group tended to have the higher speed scores, as was to be expected, an equally large number achieved speed scores as low as, or lower than, those of the older individuals.

An interesting corollary result was that those of the group whose ages were 38 or above failed to achieve speed scores above 116, whereas the group under 38 had achieved much higher scores, with no appreciable difference between the

youngest and the oldest within the age range 18 through 38. In the normal adult group were 40 individuals whose ages ranged from 38 through 62. The mean speed score for this age division was 98.5, and the mean accuracy score, 96.5.

In contrast to these scores were those obtained for the group whose ages were 37 or less. The mean speed score for the younger age group was 103.3, reliably different beyond the one per cent level of confidence from that of the older age division. The mean accuracy score for the younger age division was 97.9, which did not differ reliably from the mean accuracy score of the older group. Because of the small number of cases, no specific conclusions could be drawn from the tendencies when the cases were divided by decades. However, there did seem to be a tendency in each decade for the low speed scores to be balanced with an equally great number of cases in which the speed scores were well above 100. In the case of the accuracy scores, one was struck with the much narrower range of the scores. Few made phenomenally low or high scores; most of the scores tended to cluster close to the mean. It can be seen, therefore, that the results of this study of the speed and accuracy scores of the Kuhlmann corroborate the findings of previous studies of speed and power, in that the mean speed score of the older age group differed reliably from that of the younger age division, while the mean accuracy score failed to show any statistically reliable differentiation between the younger and older age groups.

There are also statements in the literature to the effect that speed declines faster than power (10, 37, 38, 39). No direct comparison between the scores could be made since they are in unequal units. It was established that the mean accuracy score is .55 of the maximal points possible, whereas the mean speed score has dropped to .32, a difference which is statistically reliable between the .02 and the .05 level of confidence. However, the mean speed score of the older age group did not begin to approximate the decline observed in the pathological cases, and a much greater discrepancy occurred between the speed and accuracy scores in the pathological groups, especially in those with organic involvement (11). It was interesting to find that the results of individually administered tests correlated so closely with group test results, since the motivation of individuals can differ so markedly from that of groups.

An attempt was made to see what effect vocational placement had on the speed and accuracy scores. Unfortunately, by chance, most of the group fell into one vocational level and hence no conclusions could be made from the few cases found in the other categories.

In the same way as the relationship between speed and age had been worked out, accuracy and age for the normal adult group were correlated. The correlation was $.008 \pm .063$, and the correlation ratio was $.337 \pm .056$, thus establishing a significant trend toward non-linearity. The youngest and the oldest in the group tended to have the lowest accuracy scores.

E. The Findings on the Variability Score

Another score which Kuhlmann hoped would be useful as a diagnostic measure, in addition to the speed and accuracy scores, was the variability score. In order to discover just what signifi-

cance this measure might have, many relationships were investigated such as: variability and IQ; variability and speed; variability and accuracy; variability and the difference between the Kuhlmann and the Stanford-Binet scores; variability in relation to age and sex; and the effect of the number of years the patient had been out of school upon the variability scores. An attempt was made to discover possible differences in variability scores between the patients who had undergone brain surgery and those with organic involvement who had not been operated, and also, to investigate any consistent differences in this measure which might appear between these patients and those with emotional disturbance. Also, the organic group was divided into those patients in which the pathology had progressed rapidly and those in which the pathology had progressed slowly, in order to determine what difference, if any, this factor would produce in the test results. Most of these attempts proved to be fruitless.

Chapter VI

IMPLICATIONS

A. Diagnostic Significance of IQ Differences between the Stanford-Binet and the Kuhlmann Tests

In addition to their value as measures of intelligence, the comparative scores obtained on the two tests and the speed, accuracy, and variability with which the individual works may furnish valuable diagnostic clues. In the University of Chicago Clinics the use of both tests on the same patients has been considered very helpful.

Of course, as was to be expected, the possible differences between the two tests in the feebleminded group were limited because of the narrow range of IQ and limited performance. The Stanford-Binet and Kuhlmann scores of more than half of the feebleminded individuals differed not more than 9 points, and the mean difference for the entire group, without regard to sign, was ± 7.56. Similarly, almost half of the student group varied not more than 9 points from one test to the other, with a mean difference of ± 11.53. In the normal adult group, only one-fourth varied not more than 9 points, while the mean difference of ± 16.84 was definitely higher than in the student and feebleminded groups.

However, a large difference in IQ points was found more frequently in the pathological cases. In the convulsive and organic groups the range of difference in IQ points was the same. However, of the pathological groups, the organic group stood out as having the least agreement between the two types of tests. The adult functional cases followed the pattern established by the organic pathological cases much more closely than did the student functional cases; in fact, the student functional group was surprisingly similar to the non-pathological student and adult groups.

In general, a difference in IQ points between the Stanford-Binet and the Kuhlmann tests of 17 or more points suggests the need for further investigation of emotional or organic disturbance, since the mean differences for the pathological cases were found to be more than 17 points. The single exception to this statement in the pathological cases was found among the students with emotional disturbance. This emphasizes the fact that test results of students should be kept distinct from test results of adults in general.

The direction of the difference must also be kept in mind, for the students and the normal adults tended more frequently to have higher Kuhlmann scores than did the other groups. Although the Kuhlmann was often higher than the Stanford-Binet among the students, and occasionally among the normal adults, it was almost invariably lower than the Stanford-Binet among the pathological groups.

A special word should be said about the feebleminded group with regard to IQ differences. The direction of the IQ differences was not as consistent as it was in the pathological cases, and in this particular group the direction of the difference was found to be less predictable among the women than among the men. While 61.5 per cent of the group showed an IQ difference of not more than 9 points established by the Revised Stanford retests, 9 points may be very

important around the commitment level and may call for the use and careful evaluation of more than a single measure. If the difference is great, it suggests the necessity for a review of the diagnosis with a consideration of other possible factors besides feeblemindedness.

B. Diagnostic Significance of the Speed, Accuracy, and Variability Scores

The difference between the scores of the two tests is not diagnostic in itself, but additional aids for differential diagnosis can be obtained from the internal patterns as well as from the accuracy, speed, and variability scores of the Kuhlmann. The comparative level of these scores may be of value.

Among the normal adults and the students the speed score was nearly always higher than the accuracy score, and speed, especially among the students, was often very high. Among the organic pathological groups the relationship was reversed. In the student functional cases, the speed remained higher than accuracy but did not reach as high a level as that of the unselected student. Accuracy for the adult functional cases remained fairly normal, with median speed dropping below average although the mean speed score failed to show the same tendency. The organic group differed from the emotionally disturbed adults in that the accuracy rating often fell below the average level also, and the speed score was extremely low. In the convulsive group, accuracy and speed were both down, with the speed score much the lower. The convulsive and the organic groups showed very similar trends in the speed and accuracy scores. In the feebleminded, the speed and accuracy were likely to be minimal.

Although in general the rate of speed of the various clinical groups was significantly differentiated, a few exceptions did occur. The patient with convulsions could not react appreciably differently from the patient with focal lesions because of their common organic involvement. The lack of differentiation between the normal adult and emotionally disturbed group may well have reflected the fact that any unselected adult group may possess certain unstable tendencies.

As might have been expected, the median accuracy score of the pathological cases was not as reliably different from that of the non-pathological as was the speed score since they apparently sacrificed speed to maintain their former level of efficiency.

Speed and accuracy scores may be of additional help in vocational guidance, for they provide an indication of how the person approaches a task. An individual who sacrifices accuracy for speed should not be encouraged to undertake certain types of occupations, and conversely, one who strives toward perfect accuracy would be too slow to be successful under work conditions stressing speed of performance. Also, when the need for accuracy is greatly overemphasized, the question of compulsive tendencies arises, and the accuracy should no longer be considered an asset, but rather, symptomatic of emotional disturbance. A pharmacist or radio engineer is a representative of types of occupations in which the accuracy of the individual is demanded and the speed factor is of lesser importance.

The largest variability scores tended to be concentrated in the average intelligence range in all groups. In general, those with low IQ scores tended to have low variability scores, but as the

IQ's approached the average range, the variability increased noticeably, and again decreased as the superior and maximal scores were obtained. Therefore, the median variability scores of all other total groups were lower than the median variability score of the normal adult group. The student group, as well as individuals who achieved maximal or near maximal ratings, had lower variability scores than did those in the average range; similarly the feebleminded group or those individuals obtaining low IQ scores had lower variability scores.

Although the speed and accuracy scores had shown statistically reliable levels of difference between most of the well-defined groups, very few reliable relationships were brought out when the variability scores of the different groups were compared with each other. The student group was the only one which showed statistically reliable differences from a few of the other groups. The student group when compared with the normal adult group showed the only difference which was statistically reliable at the one per cent level of confidence.

Experience with this measure is not sufficiently extended to make very positive statements, but it appears that in records where individuals with IQ's in the dull average classification or with superior or maximal intelligence have large variability scores, the reasons should be investigated. There are indications that those with low IQ's, especially on the Kuhlmann, who at the same time have high variability scores, are functioning below their potential level. Emotional blocking might account for the variable performance within the record, but one should keep in mind that regression may also be present. In the organic cases the high variability coupled with low IQ may be an indication of the amount of deterioration.

It must be remembered, however, that, since the men in all groups but the feebleminded tended to have consistently higher variability scores than the women, men would be expected to show greater variability than women.

The possibilities suggested by the IQ differences, speed, accuracy, and variability must, of course, be supported by evaluation of evidence within the tests themselves, for each type of case seems to have its own internal pattern. However, development of this theme is beyond the scope of this paper.

C. Implications for Further Study

The detailed examination of these two tests has suggested both marked differences and similarities between the Revised Stanford-Binet and the Kuhlmann with respect to the general nature of the tests. This comprehensive study concerning the relationships of the tests has brought to light certain points which might be considered in working with either one or both of these tests or in adult test construction.

The two tests utilize sufficiently different approaches that when used together they can be clinically valuable. There is evidence that specific patterns on each test exist within the various groups, which are sufficiently distinct to differentiate them. Work has been done along this line with the Stanford-Binet (43) but very little exists to date for the Kuhlmann. Such patterns, once established and used with those available from other tests, should provide valuable suggestions for further improvement of these or similar diagnostic tests.

The re-standardization of the Kuhlmann scale is contemplated for the very near future. And such standardization should take into consideration some of the points brought out in this study. As has been noted, the speed score is a very valuable measure. However, in order to place it on a more solid foundation, several changes might be considered. As it now stands, the time limits are much too generous for the very bright individual. The speed score might be made more effective first by providing much shorter time limits than now exist. Another approach might be to use the same time limits

but to add more items. Both of these would involve new standardizations. The most simplified method, the one suggested in the present study, is the cutting off of extra credits which obviously have been earned because of unusual dexterity and, in certain cases, familiarity with test procedure. In the 1933 revision, Kuhlmann used the median of the rights divided by the time scores (R/T) instead of giving credit for each successive test. If this material were still available, it would be desirable to find out whether in this way the extremely high IQ's which now occur in the maximal or very superior population might be eliminated. If the test were to be re-standardized completely, it would also be well to change some of the weak points in the individual test items which now are obvious.

Thus, results of this study have the following implications for future work with the Heinis growth curve, the Kuhlmann scale, and general test construction. Although the introduction of the Heinis growh curve to the Kuhlmann scale has proved of value, the present curve should be checked by further work. The upper end of the scale is especially weak and needs revision.

When the Kuhlmann scale is re-standardized, several changes should be considered. It should be determined whether shorter time limits, additional items, the suggested cutting of scores, or the use of medians would prove to be of the most satisfactory method of reducing phenomenally high scores which now result from the speed factor alone.

The test results on the normal student population as well as on those of the emotionally disturbed students have demonstrated outstandingly that tests standardized on the late teen age population or those who are still attending school are not suitable for the older age groups whose judgments and values have matured. Therefore, it would seem imperative that in the future, test standardization populations be classified much more strenuously from the viewpoint of age and emotional adjustment than has been attempted in the past.

It has frequently been stated, and should be re-stated, that our attempts to measure intelligence have been concerned too much with the complexity of the surface aspects and have ignored the depths. This is true to a large extent of both of the two tests considered here, particularly at certain levels, and of other tests developed for adults.

Certainly, future attempts to measure intelligence should be directed toward the development of some new ideas which may make it possible to evaluate a person's creativity and depth of intelligence, rather than what we now include in our concept of measured intelligence. Verbal recognition and reproduction now receive the greatest weight. The new test concept developed by Jastak would seem to be an answer to the chief criticism of the medical clinician that psychological test results fail to differentiate adequately those individuals whose reality contacts are greatly affected from those whose intellectual functioning remains at or near his previous maximal level.

Chapter VII

SUMMARY AND CONCLUSIONS

A. A Summary of the Agreement between the Hypotheses and the Results

1. Primary Hypotheses

A. There should be a high degree of correspondence between the Revised Stanford-Binet and the Kuhlmann on non-pathological groups, and less so for the pathological.

B. The amount of agreement would decrease in proportion to the severity of the pathological involvement; therefore, the test results of patients with organic involvement should show less agreement between the two tests than those diagnosed as having functional disorders.

C. If the clinical divisions have been properly set up, one would expect that the following classifications would be reliably differentiated from each other by a comparison of the IQ's on the two tests:

1. Normal adults ($N = 251$)

2. Students ($N = 204$)

1. Results

A. The correlations between the two tests were high for all groups, both non-pathological and pathological. The means of the two tests showed a consistent statistically reliable discrepancy in all groups.

B. The amount of agreement decreased from the functional to the organic disorders.

C. The diagnostic classifications were reliably differentiated from each other with the exception of the two organic pathological groups.

1. The normal adults differed reliably in agreement between the Revised Stanford-Binet and the Kuhlmann from all other classifications, except from the functional student group.

2. The students differed reliably in agreement between the Revised Stanford-Binet and the Kuhlmann from all other groups, except the functional student group. On the basis of differences in correlations the student group alone was reliably different from the other groups except the two organic categories.

3. Feebleminded adults ($N = 91$)

4. Patients with convulsive states ($N = 51$)

5. Patients with organic disease of the brain ($N = 43$)

6. Patients with emotional disturbance ($N = 113$)

   a. Student ($N = 28$)

   b. Adult ($N = 85$)

   c. Neurotic ($N = 66$)

D. It was felt that a comparison of the results of the Revised Stanford-Binet with the Kuhlmann would be of more diagnostic value than the results of either test alone.

3. The feebleminded were reliably differentiated in amount of agreement between Revised Stanford-Binet and Kuhlmann IQ's from all other groups.

4. Patients with convulsive states were reliably differentiated from others with the exception of the pathological groups; however, the student division of the functional group was deviant from the other pathological categories.

5. Patients with organic disease of the brain differed from the other groups in exactly the same manner as had those with convulsive states.

6. Patients with emotional disturbance differed reliably as a group from all other classifications in the same way as had the organic groups. There was no statistically reliable differentiation within its adult subdivisions.

   a. The student functional group resembled the normal adults and unselected students rather than the pathological division in amount of agreement between Stanford-Binet and Kuhlmann IQ's.

   b. The adult functional cases were identical with the main group, as was expected, for they formed the main portion of that group.

   c. The neurotic adult group followed the pattern established by the entire functional group.

D. Either the Revised Stanford-Binet or the Kuhlmann differentiated as reliably between the classifications as had the two tests used together. Identical groups were differentiated by the two tests. The tests failed to differentiate between the normal adults and functional groups, as well as between the organic groups.

E. One would not expect reliable sex differences within the groups.

2. *Secondary Hypotheses*

A. The speed and accuracy scores of the Kuhlmann should each differentiate the pathological from the non-pathological groups, the organic from the functional, and the student from the non-student.

1. One would predict that the pathological groups would drop considerably below the norms reached by the normal adult group.

2. One would predict that the patients diagnosed as having organic pathology would have lower scores than the patients with functional disorders.

3. One would expect that the scores of the student group should be definitely above average, and hence higher than those of any other group.

E. No reliable sex differences within the groups appeared.

2. *Results*

A. The speed of the normal adult was differentiated reliably from the speed of all other groups except the emotionally disturbed group and its subdivisions, while the speed and accuracy of the students was differentiated reliably from all other groups. The accuracy of the normal adult group was reliably differentiated from all but the adult emotionally disturbed group. The speed and accuracy scores did not differentiate the two organic groups. The entire functional group differed in speed from all but the normal adults, and with respect to accuracy there were no exceptions.

1. The median speed dropped below 100 for all the pathological groups except the emotionally disturbed students; the median accuracy score for the organic pathological groups dropped as did the speed, but that of the functional group remained above the norm.

2. Speed scores for both dropped, but those of the organic division reached nearly minimal levels; accuracy scores of the organic group were also low whereas those of the functional group remained relatively unaffected. The functional student group again was the exception for its speed was above norms.

3. The scores of the student group being the only scores above average automatically were the highest obtained.

4. One would expect the scores of the feebleminded group to be lower than the scores of any other group.
5. One would predict that the accuracy scores, in general, would not be affected to the same extent as the speed scores.
6. On the Kuhlmann, one would predict that the speed of the normal individual is reduced with age while the power factor remains unaffected, and that the results for individual tests would not approximate those found for group tests.
   a. The speed and accuracy scores of the older age group should not be as grossly affected as those scores of the pathological groups.
   b. The invasion of pathology should reduce power and speed to a much greater extent than the factor of age.

C. The variability score should prove to be a significant diagnostic measure so that it can be used in differential diagnosis.

4. Roughly, only one-fourth of the feebleminded group worked with sufficient speed and accuracy to reach scorable limits.
5. The accuracy score was not affected in either the organic or the functional pathological groups to the same extent as the speed score.
6. The speed score of those 38 years and over was found to be reliably different from those who were below 38 years in age. The accuracy score, though affected, did not decline as rapidly as did the speed score.
   a. The accuracy score of the older age group was not reduced at all when compared with the total normal adult group. The speed score was somewhat reduced, but not to as great an extent as in any of the pathological groups.
   b. The speed and accuracy scores of the organic groups were definitely below the levels achieved by the older age group; however, the accuracy of the emotionally disturbed group was above that of the older age group as well as the total normal adult group.

C. The variability score did not bring out the same reliable differences between the diagnostic groups; hence at present, it cannot be of help in differential diagnosis.

B. Conclusions

The relationship between the two tests, the Revised Stanford-Binet and the Kuhlmann Tests of Mental Development, was found to differ according to the nature of the group involved. In general, a consistent though not always reliable relationship between the tests was established. The size and the direction of the differences were dependent

not only upon the degree of brightness of the individual, but also upon the diagnostic classification.

Because of the *consistent* differences between scores on the two tests, the correlations obtained were high for all groups and, therefore, this method of comparison failed to bring out any statistically reliable differences between most of the clinical classifications. In all groups the means differed reliably between the two tests.

The difference in the level of performance of the same individual on the Revised Stanford-Binet and the Kuhlmann may be due to several reasons. One fact may be the fundamental difference in the construction of the two tests. One would surmise that in the case of the superior individual the Kuhlmann score would be higher than the Revised Stanford-Binet (cf. Table 4). The Revised Stanford-Binet seems a poor measure at this level because of its low ceiling, lack of variety of items, presentation of the same type of test at successive year levels, and its questionable use of the vocabulary test beyond the average adult level. The general tendency for the pathological cases in this study to score higher on the Stanford-Binet, as well as the slight changes observed in pre- and post-operative cases, would demonstrate the general insensitivity of this measure. However, factors in favor of the Revised Stanford-Binet are: (a) the vocabulary test provides a rough estimate of the original level and thus the amount of possible deterioration; (b) at the dull and average levels it does provide an excellent diagnostic measure and is relatively free from the above enumerated weaknesses.

On the other hand, the Kuhlmann test is criticized for its complicated scoring and its unsuitability as a measure of intelligence in cases of reading disability. The fact that the Kuhlmann test reflects, as demonstrated by the much lower means, a markedly lower level in the performance of most pathological cases may indicate it to be a much more sensitive instrument for detecting lowered efficiency. Such a point in its construction as the extremely short time for the presentation of numerous items forces the individual to shift quickly from one task to another. Consequently, persons who lack flexibility or whose ability to shift their set has been reduced, perform less ably than those who retain this capacity. The Kuhlmann, although presented in simple language, does not permit dependency on previous learning, and, therefore, one cannot depend upon the use of ready phrases or retained knowledge which is no longer incorporated meaningfully. Other advantages of the Kuhlmann are: the high ceiling allowing the evaluation of superior individuals, and its use of few arithmetical processes.

Since the difference between the means of the groups did not take into consideration the range of individuals' differences in scores on the two tests, the absolute difference (without regard to sign) in IQ points between the two measures for individuals was calculated. The size of the difference in IQ points ranked the different groups in the following order (from low to high): the feebleminded, the students,[6] the normal adults, those with functional disorders (and its adult subdivisions), those with convulsive states, and finally those with organic disease of the brain. None of those with emotional disturbance was markedly deteriorated. While most of the individuals with organic pathology had suffered severe involvement, the amount of agreement between the two tests generally decreased in proportion to the severity of the pathology present. The small size in the difference in IQ points observed for the feebleminded group was due to the limitation by classification of the ceiling possible for such a group.

Marked difference of the adult group from other groups[7] appeared. The standard deviation of this group on the Revised Stanford-Binet was found to be 20.8, which substantiates the prediction

[6] The functional student group followed the general student group. Also, the emotionally disturbed student group differed from the other pathological groups in that the size of the difference in IQ points resembled the difference found in the non-pathological groups.

[7] Again, the emotionally disturbed student group was closer to the non-functional groups than the functional subdivisions.

made in the more recent studies that the standard deviations in adult groups might be expected to be greater than those found in groups of children. As was to be expected from the wide range possible on the Kuhlmann, the standard deviation was higher than that of the Revised Stanford-Binet. Diagnostically, any difference in IQ points of 17 or more between the Stanford-Binet and the Kuhlmann tests indicates the necessity for considering the possibility of emotional or organic disturbance.

The group of feebleminded adults differed significantly from all other groups, thus justifying their separation from the normal adult group or the deteriorated cases of the pathological divisions. The range of difference between the two test scores for the feebleminded group was about half that found for the normal adult group and much less than that of any of the pathological groups. Because of this general tendency, the warning made by numerous investigators should be reiterated, that sizeable differences between the two tests at the upper limits of the feebleminded classification should suggest the possibility of a diagnosis other than feeblemindedness.

The two groups of patients with organic diagnoses failed to be reliably differentiated from each other as well as from the emotionally disturbed group according to the difference of the IQ points between the two tests.

The decision to separate the entire emotionally disturbed group into the various subdivisions was found as a whole to be sound. This was particularly advisable in the separation of the students from the adults.

The division of the adult emotionally disturbed group into its neurotic and psychotic components did not prove to be as striking as the student and non-student subdivision. The psychotic students, in contrast to the non-psychotics, tended to show the trend established by the pathological groups in general. The small number of the psychotic group alone in itself makes any of the findings most questionable.

Both the revised Stanford-Binet and the Kuhlmann test differentiate the diagnostic groups in an identical manner, despite the markedly different approach of the two tests and the observedly different level of achievement on them. However, the use of the two tests in combination is a superior method to the use of either test alone in differentiating emotionally disturbed from normal individuals. Even though identical differential diagnostic capacity for certain pathologies was brought out for the Revised Stanford-Binet and the Kuhlmann individually, the Kuhlmann actually is the more diagnostic measure when one considers the additional weight provided by the speed and accuracy scores.

As expected, significant sex differences between the groups failed to appear.

The speed and accuracy scores were found to yield statistically reliable differences between the non-pathological and pathological cases, and can be used to assist in making a differential diagnosis.

In the same manner, the mean speed and accuracy scores of the organic groups are to a statistically reliable degree different from the mean scores of a functional group. The close-to-the-minimal speed scores of the patients with organic pathology reflect the impact of the organic process to a much greater extent than the low speed scores of the adult functional cases show the influence of the emotional disturbance. The accuracy of the patients with organic involvement is likewise reduced, but not to as great an extent as the speed. Interestingly enough, the accuracy scores of the patients with emotional disturbance tend to remain near the average level.

The reliable differentiation brought out between the student and non-student groups in IQ results is carried out in regard to the speed and accuracy scores. The unselected student group is reliably superior to the normal adult group; but this is not true of the students with emotional disturbance. Even severe emotional disturbance in the case of students, however, is usually not sufficient to drop their speed to the level of adult functional cases. This speed factor may be responsible for the marked deviation of the student functional group.

The accuracy scores of the cases with severe pathological involvement, although affected, are not lowered to the same extent as are the speed scores.

The predictions in the literature that a remarkable difference exists between the speed and power factors in the older age group were upheld in a comparison of the speed and accuracy scores of individually administered Kuhlmann tests. A large speed-accuracy difference may suggest definite pathology, in addition to the gradual process of aging.

The variability score on the Kuhlmann test, in general, cannot yet be regarded as a useful measure in differential diagnosis. It does not reflect, as might have been supposed, the amount of disturbance of the speed or the accuracy of performance. Rather, the size of the variability score is related to the IQ level. Those of average intelligence are much more variable in their performance.

An attempt was made to obtain the then unpublished Wechsler-Bellevue Intelligence Scale, but it was not available. It is hoped that a more comprehensive comparison in which results from this test are also involved can be made.

In conclusion, results of this study have the following implications for general test construction. The differentiation of the various clinical groups on the basis of the patterns found on the Stanford-Binet and the Kuhlmann would indicate that not only age but recency of school attendance should be considered in forming groups for test standardizations. Also, the emotional adjustment of the individual needs more consideration in the selection of subjects than has been the case in the past. It appears that there is a direct relationship between the type and the severity of the disturbance and its effect upon the test results.

The knowledge which has accrued from the study of behavior has not been incorporated into the measurement of intelligence, and this lack of integration of knowledge no doubt hinders progress in test construction.

BIBLIOGRAPHY

1. ALEXANDER, F. The neurotic character. *Internat. J. Psychoanal.*, 1930, 11, Part 3, 292-311.
2. ANDERSON, E. E., *et al.* Wilson College studies in psychology: a comparison of the Wechsler-Bellevue, Revised Stanford-Binet and American Council of Education Tests at the college level. *J. Psychol.*, 1942, **14**, 317-326.
3. BARANYAI, E. A. Superior intelligence and its examination. Abstracted in *Psychol. Abstr.*, 1938, **12**, 302.
4. BECKMAN, R. O. A new scale for gauging occupational rank. *Personnel J.*, 1934, **13**, 225-233.
5. BENNETT, G. K., SEASHORE, H. G., & WESMAN, A. G. *Differential aptitude tests.* New York: Psychological Corp., 1947.
6. BINGHAM, W. V. *Aptitudes and aptitude testing.* New York: Harper, 1937.
7. CARLTON, T. A comparison of the Revised Stanford-Binet, Form L, with the Kuhlmann Tests of Mental Development. *J. genet. Psychol.*, 1942, **61**, first half, 47-54.
8. CARLTON, T. A comparison of the Revised Stanford-Binet, Form L, with the Kuhlmann Tests of Mental Development: II. Performances of mentally defective children. *J. appl. Psychol.*, 1942, **26**, 159-167.
9. CATTELL, R. B. A Culture Free Intelligence Test I. *J. educ. Psychol.*, 1940, **31**, 161-180.
10. CATTELL, R. B. The measurement of adult intelligence. *Psychol. Bull.*, 1943, **40**, 153-193.
11. CONRAD, H. S. The measurement of adult intelligence and the requisites for a general intelligence test. *J. soc. Psychol.*, 1931, **2**, 72-86.
12. CONRAD, H. S., JONES, H. E., and HSAIO, H. H. Sex differences in mental growth and decline. *J. educ. Psychol.*, 1933, **24**, 61-69.
13. GERSTEIN, S. A suggestion for a revision of the Revised Stanford-Binet Examination applied to adults. *J. Psychol.*, 1941, **12**, 225-226.
14. GREENE, E. B. *Measurements of human behavior.* New York: Odyssey Press, 1941.
15. HALLORAN, R. D., and YAKOVLEV, P. I. *Collected lectures of the seventh postgraduate seminar in neurology and psychiatry, second semester: general psychiatry.* Waltham, Mass.: 1942.
16. HALSTEAD, W. C. A power factor (P) in general intelligence: the effect of brain injuries. *J. Psychol.*, 1945, **20**, 225-226.
17. HARTSON, L. D. Does college training influence test intelligence? *J. educ. Psychol.*, 1936, **27**, 481-491.
18. HEBB, D. O. The effect of early and late brain injury upon test scores, and the nature of normal adult intelligence. *Proc. Amer. Phil. Soc.*, 1942, **85**, 275-292.
19. HEBB, D. O., and MORTON, N. W. Note on the measurement of adult intelligence. *J. gen. Psychol.*, 1944, **30**, second half, 217-223.
20. HEINIS, H. A. A Personal Constant. *J. educ. Psychol.*, 1926, **17**, 163-186.
21. HEINIS, H. A. La loi du développement mental. *Arch. de Psychologie*, 1924, **19**, 97.
22. HESTON, J. C., and CANNELL, C. F. A note on the relation between age and performance of adult subjects on four familiar psychometric tests. *J. appl. Psychol.*, 1941, **25**, 415-419.
23. HILDEN, A. H. A comparison of the Intelligence Quotient and the Heinis Personal Constant. *J. appl. Psychol.*, 1933, **17**, 355-375.
24. JASTAK, J., and VIK, ESTHER STUBBS. The objective measurement of reality perceptions in dementia praecox. *Delaware State Med. J.*, 1947, **19**, 98-102.
25. JONES, H. E., and CONRAD, H. S. The growth and decline of intelligence. *Genet. Psychol. Monogr.*, 1933, **13**, 223-298.
26. KUHLMANN, F. *A handbook of mental tests.* Baltimore: Warwick and York, 1922.
27. KUHLMANN, F. *Tests of mental development.* Stillwater, Minn.: Minnesota State Board of Control, the Division of Research, preliminary edition, 1933.
28. KUHLMANN, F. *Tests of mental development.* Minneapolis: Educational Test Bureau, 1939.
29. KUHLMANN, F., and ANDERSON, ROSE G. *Kuhlmann-Anderson intelligence tests.* Minneapolis: Educational Test Bureau, Second edition, 1930.
30. KUHLMANN, F., and ODOROFF, M. E. Verification of the Heinis Mental Growth Curve on results with the Stanford-Binet tests. *J. Psychol.*, 1942, **13**, 356.
31. LAWTON, G. Mental abilities at senescence. *J. appl. Psychol.*, 1938, **22**, 607-619.
32. LAWTON, G., FOLSOM, J. K., COHN, A., LORGE, I., SCHILDER, P., and KAUFMAN, M. R. Old age and aging. *Amer. J. Orthopsychiat.*, 1940, **10**, 27-88.
33. LORGE, I. Intellectual changes during maturity and old age. *Rev. educ. Res.*, 1941, **11**, 553-561.
34. LORGE, I. Intellectual changes during ma-

turity and old age. *Rev. educ. Res.*, 1944, **14**, 438-445.

35. LORGE, I. The influence of the test upon the nature of mental decline as a function of age. *J. educ. Psychol.*, 1936, **27**, 100-110.
36. MCFARLAND, R. A. An experimental study of the relationship between speed and mental abilities. *J. gen. Psychol.*, 1930, **3**, 67-97.
37. MILES, CATHERINE C. Influence of speed and age on the intelligence scores of adults. *J. gen. Psychol.*, 1934, **10**, 208-210.
38. MILES, CATHERINE C., and MILES, W. R. The correlation of intelligence scores and chronological age from early to late maturity. *Amer. J. Psychol.*, 1932, **44**, 44-78.
39. MILES, W. R. Age and human ability. *Psychol. Rev.*, 1933, **40**, 99-123.
40. MITCHELL, MILDRED B. Irregularities of university students on the Revised Stanford-Binet. *J. educ. Psychol.*, 1941, **32**, 513-522.
41. MITCHELL, MILDRED B. The Revised Stanford-Binet for adults. *J. educ. Res.*, 1941, **34**, 521.
42. MITCHELL, MILDRED B. The Revised Stanford-Binet for university students. *J. educ. Res.*, 1943, **36**, 510-511.
43. MYERS, C. R., and GIFFORD, ELIZABETH B. Measuring abnormal patterns on the Revised Stanford-Binet Scale (Form L). *J. Mental Sci.*, 1943, **89**, No. 374 (New Series No. 338), 92-101.
44. PRESSEY, S. L., and ROBINSON, F. P. *Psychology and the new education.* New York: Harper, 1944.
45. RAPAPORT, D. *Diagnostic psychological testing.* Chicago: Year Book Publishers, 1946. Vol. I.
46. RAUTMAN, A. L. Relative difficulty of test items of the Revised Stanford-Binet. An analysis of records from a low intelligence group. *J. exp. Educ.*, 1942, **10**, No. 3, 183-194.
47. RILEY, G. A comparison of the P.C. and IQ. *Psychol. Clinic*, 1930, **18**, 261-265.
48. RILEY, G. Stanford-Binet "indicators" of mechanical ability. *Psychol. Clinic*, 1930, **18**, 128-132.
49. SCOTTISH COUNCIL FOR RESEARCH IN EDUCATION. *The intelligence of school children: a national survey of an age group.* London: Univ. of London Press, 1933.
50. SHAKOW, D. Intellectual changes with age. *Public Health Reports, Supplement No. 168*, 1941, 43-52.
51. SOLOMON, H. C., and YAKOVLEV, P. I. *Manual of military neuropsychiatry.* Philadelphia: Saunders, 1944. Pp. 128-561.
52. STODDARD, G. D. On the meaning of intelligence. *Psychol. Rev.*, 1941, **48**, 250-260.
53. STRAUSS, A. A. Effects of exogenous and endogenous factors in the organism as a whole in mentally deficient children. *Proc. 3rd Bienn. Meeting Soc. Res. Child Develpm.*, 1939, 93-95.
54. STRAUSS, A. A., and LEHTINEN, LAURA E. *The psychopathology and the education of the brain-injured child.* New York: Grune and Stratton, 1947.
55. SWARD, K. Age and mental ability in superior men. *Amer. J. Psychol.*, 1945, **58**, 443-479.
56. TERMAN, L. M., and MERRILL, MAUDE A. *Measuring intelligence.* New York: Houghton Mifflin, 1937.
57. THOMPSON, W. S. *Population problems.* New York: McGraw-Hill, Third edition, 1942.
58. WECHSLER, D. *The measurement of adult intelligence.* Baltimore: Williams and Wilkins, 1939.
59. WERTHEIMER, M. *Productive thinking.* New York: Harper, 1945.
60. YAGER, J. L. Comparison of the results of the Kuhlmann Tests of Mental Development and the Revised Stanford-Binet, Form L, in normal children. Paper read before the meeting of the Midwestern Psychological Association, Chicago, Illinois, May 3, 1947.